

# VU Research Portal

## Evaluating Ontology-Alignment Techniques

van Hage, W.R.

2008

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

van Hage, W. R. (2008). *Evaluating Ontology-Alignment Techniques*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# SUMMARY

This thesis is about ontology alignment, the linking of structured vocabularies.

The World Wide Web has changed the way information flows; it has democratized information access. Anybody can provide information to anyone from his own computer. This has made it possible to easily draw information from various sources, for example, libraries, weather services, or bookstores, without the effort of physically travel, by visiting their websites. Due to this new convenience it is now feasible to combine multiple collections for tasks like federated search.

In practice, the combination of collections is complicated by the fact that objects in different collections are often described using different vocabularies, each with its own view on the domain. These views can differ in many aspects. They can use different languages, they can use different words for the same things (*e.g.* different jargon), they can describe things from different points of view (*e.g.* insulin as a hormone or as a protein), and they can disagree on the meaning of things (*e.g.* when they have different experiences or different idiom). These differences introduce a semantic gap between the collections that will have to be bridged to combine the collections without conflicts. In the past this was commonly done by merging the vocabularies. In this thesis we investigate the possibility to use ontology alignment to bridge the gap.

In the past years, the research field of ontology alignment has rapidly developed into maturity. Many different alignment techniques have been developed. Many organizations, especially libraries and museums, have undertaken alignment projects to interconnect the ontologies (usually simple vocabularies, thesauri, or classification schemes) they use to describe their collections. Since 2004 the Ontology Alignment Evaluation Initiative (OAEI) has hosted an increasing number of widely varying alignment tasks ranging from the alignment of web directory structures to rich medical ontologies, to compare the quality of alignment techniques in various settings.

The work described in this thesis contributions to the field of ontology alignment in the following ways.

We extend the range of ontology alignment techniques. Most of the existing alignment techniques are designed to find equivalence relations using only the ontologies as input. We introduce complementary methods to find subclass and part-whole relations from textual background knowledge, for example, web pages or dictionaries.

We set up comparative evaluation tasks for alignment techniques. Most alignment projects have to do with library collections indexed with large thesauri. We introduce two tasks to the OAEI, the food and environment tasks, to measure the quality of alignment techniques on matching the thesauri of the United States Department of Agriculture (the NAL Agricultural Thesaurus), the Food and Agriculture Organization of the United Nations (AGROVOC), and the European Environment Agency (GEMET). We measure and compare the performance of the seven alignment systems that participate in these tasks.

We investigate factors that influence how well the various techniques implemented by the systems perform.

One aspect in which the thesauri of the OAEI food and environment tasks are typical is that they are large. Hence, the alignments between them are also large. We describe a method for sample-based evaluation of ontology alignment to make the comparison of alignment techniques on large ontologies feasible.

In the OAEI automatic ontology alignment systems are compared to each other. This allows us to conclude which techniques outperform others in certain cases. However, it does not tell us how these techniques relate to human experts. Most of the current alignment work is carried out by human experts. To conclude how the automatic techniques would fare in practice we analyze the difference between a manually-created alignment (the alignment between the AGROVOC thesaurus and the German national library's Schlagwort-normdatei) to the automatically-created alignments of the OAEI food task.

The evaluation tasks of the OAEI do not take into account which part of the alignment is most useful in practice. Most tasks assume that every mapping is equally valuable. In practice, this is not true. The typical case is not equal to the average case. We describe a method to draw samples to represent typical usage. We apply this method to the OAEI food task to complement its average case estimates.

The research questions addressed in this thesis question the quality and applicability of ontology alignment in practice. We find that ontology alignment can already be used effectively to partially automate human labor, but that it is far from a solved problem. Furthermore, we find that task-specific evaluation techniques are the most suitable for the evaluation of ontology alignment, because the perceived quality of an alignment greatly depends on the nature of the domain and the type of the problem one tries to solve.

We conclude this thesis with advice to prospective users of alignment techniques on how to set up an alignment initiative, and to fellow researchers on promising directions of future work.